Classification of Brain Tumours from MR Spectra: The INTERPRET Collaboration and its Outcomes

Short Title: Robust classification of brain tumours from MR spectra

Margarida Julià-Sapé [1,2,3] *, John R. Griffiths [4] *#, A. Rosemary Tate [5], Franklyn A. Howe [6], Dionisio Acosta [7], Geert Postma [8], Joshua Underwood [9], Carles Majós [10,1], Carles Arús [2,1,3] #

*Equal contribution

#Corresponding author

[1]: Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola del Vallès, Spain.

[2]: Departament de Bioquímica i Biologia Molecular, Unitat de Bioquímica de Biociències, Edifici Cs, Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain.

[3]: Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain.

[4]: Cancer Reasearch UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, United Kingdom.

[5]: School of Informatics. University of Sussex. Falmer Brighton BN1 9QJ, UK.

[6]: Cardiovascular and Cell Sciences Research Institute, St George's, University of London, Cranmer Terrace, London, SW17 0RE, United Kingdom.

[7]: CHIME, University College London. The Farr Institute of Health Informatics Research. 222 Euston Road. London NW1 2DA, United Kingdom.

[8]: Radboud University Nijmegen, Institute for Molecules and Materials, Analytical Chemistry. Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

[9]: London Knowledge Lab, Institute of Education. 23-29 Emerald Street. London WC1N 3QS, United Kingdom.

[10]: Institut de Diagnòstic per la Imatge (IDI), CSU de Bellvitge, Autovia de Castelldefels km 2.7, 08907, L'Hospitalet de Llobregat, Barcelona, Spain

KEYWORDS

human, interpret, brain, tumour, magnetic resonance spectroscopy, classifier, database, decision-support system

ABBREVIATIONS

A2: Astrocytoma WHO grade II.
A3: Astrocytoma WHO grade III.
AB: Abscess.
AGG: Aggressive tumour (GB and ME).
ANN: Artificial Neural Network.
AUC: Area Under the Curve.
CCA: Canonical correlation analysis.
CDVC: Clinical Data Validation Committee.
CSF: Cerebrospinal fluid.
DMS: Data manipulation software.
DSS: Decision-support system.
eTumour: Web Accessible MR Decision Support System for Brain Tumour Diagnosis and Prognosis, Incorporating in vivo and ex vivo Genomic and Metabolomic Data.
GB: Glioblastoma.
iDB: INTERPRET database.
ICA: Independent Component Analysis
INTERPRET: International network for Pattern Recognition of Tumours Using Magnetic Resonance.
jDMS: Java data manipulation software.
jMRUI: Java MRUI.
KNN: K-nearest neigbour algorithm.
LDA: Linear discriminant analysis.
LGG: Low grade glial tumour.
LY: Lymphoma.
ME: Metastasis.
mI/Gly: m-Inositol glycine ratio.
MN: Meningioma.
MRUI: Magnetic resonance user interface.
NO: Normal brain.
OA: Oligoastrocytoma WHO grade II.
OD: Oligodendroglioma WHO grade II.
PACS: picture archiving and commnication sytem (PACS).
PCA: Principal component analysis.
PN: Primitive neuroectodermal tumour.
QC: Quality control.
QDA: quadratic discriminant analysis.
SVM: support vector machines.
TE: echo time
TR: recycling time.
viDB: validated INTERPRET database.
WHO: World Health Organisation.

ABSTRACT

The INTERPRET project was a multicentre European collaboration, carried out from 2000 to 2002, which developed a decision-support system (DSS) for helping neuroradiologists with no experience of MRS to utilise spectroscopic data for the diagnosis and grading of human brain tumours. INTERPRET gathered a large collection of MR spectra of brain tumours and pseudo-tumoural lesions from seven centres. Consensus acquisition protocols, a standard processing pipeline and strict methods for quality control of the aquired data were put in place. Particular emphasis was put on ensuring the diagnostic certainty of each case, for which all cases were evaluated by a clinical data validation committee. One outcome of the project is a database of 304 fully-validated spectra from brain tumours, pseudotumoural lesions and normal brains, along with their associated images and clinical data, which remains available to the scientific and medical community. The second is the INTERPRET DSS, which has continued to be developed and clinically evaluated after the project ended.

We also review here the results of the post-INTERPRET period. We evaluate the results of the studies with the INTERPRET database by other consortia or research groups. A summary of the clinical evaluations that have been performed on the post-INTERPRET DSS versions is also presented. Several have shown that diagnostic certainty can be improved for certain tumour types when the INTERPRET DSS is used in conjunction with conventional radiological image interpretation. About 30 papers concerned with the INTERPRET single voxel dataset have so far been published. We discuss stengths and weaknesses of the DSS and the lessons learned. Finally we speculate on how the INTERPRET concept might be carried into the future.

271 words, max 300.

THE ORIGINS OF INTERPRET

Magnetic Resonance Spectroscopy (MRS) of cancers in the human body, which has been possible for 30 years (1), provides a unique and entirely non-invasive method for detecting and quantifying metabolites within tumours. It can nowadays be performed on standard clinical MRI instruments; obtaining a matrix of MR spectra from a suspicious cerebral mass adds about 5-10 minutes to a routine diagnostic evaluation at 3T. However, despite many years of successful research, MRS is rarely used as a routine clinical method. Likely issues that have inhibited its use include a perceived requirement for the end-user to understand biochemistry and metabolism, and the need for time-consuming and expert data manipulation, especially if quantitative metabolite concentrations are required.

During the early 1990s Sian Howells and colleagues at SGUL (partner acronyms are given in table S1) demonstrated that subjective expert interpretation of MR spectra was not always necessary for tumour diagnosis, since computer-based pattern recognition methods (sometimes termed "chemometrics") could classify ^{31}P MRS spectra from animal tumour biopsies (2) and from animal tumours scanned non-invasively in vivo (3). Several studies over the next few years (4,5) demonstrated that statistical methods could be used to classify ^{1}H MR spectra of human brain tumours, indicating both the type of brain tumour (glioma, meningioma etc.) and in some cases its grade of malignancy. A highly malignant glioblastoma multiforme, for instance, has a completely different spectrum from a low-grade glioma (Figure 1). However these early brain tumour classsifiers required quantification of the peaks and consequently, they were not widely adopted.

In contrast, the chemometric classification methods developed by Howells (3,6) used "raw" spectra, with no quantification or even assignment of the peaks. If such an approach could be

applied to in vivo ^{1}H spectra of human tumours there would be no need for expert interpretation or manipulation of the spectrum, opening the way for the development of an automated method for spectral classification that could be used by radiologists or other medical practitioners with no specialist knowledge of MRS. To achieve this a large dataset of spectra from tumours with known diagnoses would be required, in order to train the classifier to recognise a spectra of a new patient undergoing diagnosis..

The initial grant application for what became the INTERPRET project was developed in the late 1990s by Rosemary Tate and John Griffiths, then at SGUL; this application responded to a call from the European Union Framework 5. However, several issues had to be addressed before the application could be completed (7).

Development of the INTERPRET collaboration

First, it was clear that a collaborative project would be needed to create a database of the size required for automated spectral analysis. Second, data analysis experts would be needed to create algorithms for classifying the spectra.

A third issue was the format for presentation of the results to the user. Given the large variety of cancer types, some very rare, it was unlikely that a prototype system could give a clear cut diagnosis. Preliminary work at Sussex University (8) showed that the best approach was to implement a decision support system (DSS) that would display the unknown case in a feature-space that demonstrated how closely it corresponded to the spectra of cases with known diagnoses. Since the intended users would probably be radiologists, images and a limited amount of clinical data should be provided with each case. All this information would need to be uploaded into the developing database over the internet and then be subject to quality control and curation. Experts in all these matters were therefore recruited to the collaboration.

Brain tumours were the obvious exemplar for several reasons. Lesions in the brain give better spectra than those from other parts of the body, because (i) the normal brain lacks the lipid deposits that give confounding peaks in tumour spectra from elsewhere in the body (although scalp lipids can be a problem for some peripheral brain tumours); (ii) the brain is not subject to significant respiratory motion; (iii) it was already known that brain tumours could be classified from their ^{1}H spectra (4,5); (iv) most brain tumours are currently diagnosed by a histopathologist, using a specimen obtained by stereotactic needle biopsy, which is a very unpleasant experience for the patient and incurs morbidity and even mortality (9-11), so a non-invasive method that could reduce the need for brain biopsy would be welcome(12); (v) several European groups were already working on brain tumour MRS, which facilitated the recruitment of collaborators.

The resulting INTERPRET consortium consisted of radiologists, computer scientists and biochemists , plus neuro surgeons and neuropathologists. The EU call required that the collaboration should include industrial partners to ensure that the diagnostic tool developed by INTERPRET would be commercially marketed, so a software company (PRAXIM) and an instrument company (Siemens) were invited to join.

The project took the form of an EU-funded collaboration (*IST*-1999-10310), from January 1st, 2000, to December 31st, 2002, and was led by Carles Arús, from the Universitat Autònoma de Barcelona in Spain. Seven clinical magnetic resonance centres in six countries participated in the prospective data acquisition (Table S1).

MR methods

The participating centres used the following 1.5T machines: GE Signa, Signa Advantage, and LX CV/i 1.5T, Philips NT and ACS NT 1.5T and Siemens Vision 1.5T. Consensus acquisition parameters for Single Voxel (SV) spectroscopy are summarised in Table S2. Before the MRS

acquisition, a basic set of MR images was acquired to ensure that the voxel was entirely located within the lesion and avoid contamination from normal adjacent brain parenchyma or oedema. When contrast was administered, MRS was performed after contrast. Two centres also performed multivoxel MRS imaging (MRSI), one using the PRESS long TE protocol and the other the STEAM short TE protocol (http://gabrmn.uab.es/interpret/mrs_data/mrs_data.html).

MRS processing pipeline

Consensus acquisition protocols were developed to minimise differences in the data format, and the post-processing algorithms and methodologies were standardised (13,14). In 2000, the DICOM standard for spectroscopy was still under development, and retrospective data came with a rich variety of spectral widths and numbers of data points (Table S2). Therefore, an automated processing pipeline, the data manipulation software (DMS), was developed for converting spectra into an "INTERPRET Canonical Format", a 512-point spectrum covering the [-2.7, 7.1] ppm range (Table S3). The canonical spectra had three different uses: (i) classifier development, (ii) display of the spectrum for quality control, (iii) allowing radiologists to input spectra into the DSS without cumbersome manual processing.

Quality control of MRS data

The quality assessment system ensured traceability of records and quality control (QC) records for both MRS and clinical data (15). Each instrument's performance was checked bimonthly and, more rigorously, annually by measurements on a specially designed phantom (15) and by taking spectra from five healthy volunteers. QC of spectra uploaded to the database was performed automatically on two indices calculated by the DMS: the signal-to-noise ratio of the processed spectrum and the linewidth of the water peak in the unsuppressed water file (Table S4). A final manual check by a committee of expert spectroscopists looked for other artefacts

such as large baseline distortions, residual water peaks or large phasing errors. All QC information is stored in the iDB as metadata.

Quality control of clinical data

This was the responsibility of the Clinical Data Validation Committee (CDVC) of practising neurosurgeons, neuroradiologists and cancer clinicians and chaired by Antony Bell, (http://gabrmn.uab.es/interpret/clinical_data/clinical_data.html)(16). They verified each case and tagged those suitable for classifier development. The main criteria were a consensus diagnosis and appropriate clinical information (i.e. age, sex, tumour location). The histological diagnoses from either needle biopsies or surgical specimens were verified by neuropathologists (Peter Wilkins, Isidre Ferrer and Pieter Wesseling) who also checked each other's diagnoses for consistency (17,18).

Databases

The final version of the INTERPRET iDB contains MRS data (SV or MV spectra) for 775 patients (Table S1). Since paraffin biopsy sections could not be obtained from all retrospective cases, a consensus histopathological diagnosis is available for 477 cases. The number of cases with at least one good quality SV short echo spectrum, acquired from the solid part of the mass in the region where the biopsy or tumor resection took place is 282. These latter cases, together with data from 22 normal volunteers, form the viDB (18), which continues to be the project's publicly available database (http://gabrmn.uab.es/interpretvalidateddb/).

Classification of MR spectra

Inevitably there were more cases from common cancers and too few cases of many rarer types to form adequate training sets. Certain cancer types were therefore aggregated together in order to have large enough groups for classifier development. The viDB contained short TE

spectra for 35 low-grade glial tumours (LGG) -comprising 22 astrocytomas (A2), 6 oligoastrocytomas (OA) and 7 oligodendrogliomas (OD) of WHO grade II- 123 aggressive tumours (AGG) -comprising 85 glioblastomas (GB) and 38 metastases (ME) - and 62 low-grade meningiomas (MN) which included meningiomas of WHO grades I and II. The rest of cases corresponded to 22 normal volunteers, 10 malignant brain lymphomas (LY), 8 abscesses (AB), and 44 cases from 17 different pathologies and/or grades (18).

The software classifiers had to categorise the spectra purely according to their metabolic profiles, uninfluenced by issues such as scanner brand, pulse sequence or TE. A preliminary study (19) developed classifiers using spectra obtained with different instruments, by STEAM or PRESS sequences, and with TEs ranging from 20 to 32 ms. None of those factors affected classification performance, provided that all the spectra had been processed in the same way and interpolated to the same number of points and sweep width.

The definitive INTERPRET classifier, a short TE classifier for the most common brain tumour types, was developed using features selected by correlation analysis(19). It was trained to distinguish three classes: LGG, AGG, and MN. The number of features was restricted according to the size of the training set (generally n/3, where n is the number in the smallest group) to avoid overfitting (20,21). The “LGG vs. AGG vs. MN” classifier gave excellent results with the independent test set (89% accuracy) and the short TE classifier was chosen over the long TE one for its slightly better performance. Development of a “GB vs. ME” classifier was also attempted; it showed reasonable results (70-80% correct classifications) when data from a single centre were used, but when it was tested with data from the other centres (19) accuracy dropped to about 62% (short TE) and 48.7% (long TE).

These INTERPRET classifiers, developed at SGUL by Rosemary Tate, used a very simple technique, linear discriminant analysis (LDA), run on commercial programs (SAS, SPSS).

In another SGUL study (22) Christophe Ladroue explored the potential of independent component analysis (ICA) to automatically extract features from brain tumour spectra that related to the underlying metabolite signals (23). Early studies at St George's (23) and in Sabine van Huffel's group from KUL, also demonstrated that for many classification tasks (e.g. those involving the LGG, AGG, and MN tumour groups) sophisticated non-linear methods were no more successful than simple linear methods for short TE (24) and long TE (25) spectra.

There were also attempts to build classifiers for the multivoxel data, initially using information contained in the variously acquired MRI images (T1, T2, Proton Density and Gadolinium Enhanced T1). Unsupervised classifiers were constructed that segmented an image and objectively identified the possible tumorous area (26). A supervised classifier (LDA) discriminated between healthy and tumour regions and also between OD and A2 in a limited set of patients (26-28).This system was capable of segmenting and identifying the volume of interest into voxels belonging to healthy tissue, cerebrospinal fluid, WHO grade II, III and IV glial tumours, unknown areas and voxels on which no decision could be made. The voxels were coloured according to these classes and an associated probability of membership to each class was provided in a prototype Decision Support System for multivoxel data (29) (Figure 2).

The Decision Support System (DSS)

The SV DSS was developed by the Human-Computer Interaction team at UOS (30). Key users (radiologists and spectroscopists, both experts and beginners in MRS) , helped to set the functional requirements of the system, and later to evaluate versions containing preliminary datasets and classifiers (31). Usability aspects tested included whether a 2D or a 3D display of cases in the classification space was more useful, and how the system was used by different user types, with a video camera recording user's interactions for further analysis. The final DSS prototype (31) was both a visual interface, displaying the cases in the database, and an

automated classifier for new cases. The left-hand panel presents the distribution of cases in the database as an interactive 2D scatterplot using the first two dimensions of the selected classifier. Cases in this space, represented by a symbol indicating the tumour type, can be selected and case data including spectra displayed and compared in the right-hand panels (Figure 3~~Figure 3~~). A spectrum from an undiagnosed case is automatically positioned in the scatterplot and users may choose to compare spectra from the new case with spectra from similarly positioned known cases or with averaged plots of the spectra of selected tumour classes. The DSS can also be used to create and explore the distribution of spectra from selected tumour groups, in interactive 2D scatterplots of spectrum intensities or intensity ratios at user chosen ppm values.

INTERPRET AFTER INTERPRET

Work on many aspects of the INTERPRET project continued after the end of its funding period.

From the INTERPRET DSS prototype to the industrial INTERPRET DSS

To turn the prototype DSS into a commercially marketable system, PRAXIM developed the "Industrial INTERPRET DSS". In 2003, PRAXIM passed its rights to SCITO, a related company, which continued that work. In the first step towards CE accreditation, the UAB team verified the traceability of all the cases in the database, obtaining a final list of 304 cases. SCITO then re-engineered the system as one of the modules in an automated client-server application called RADIONET, which was intended to provide a unified solution for the radiological examination process (http://www.scito.com/produit_radionet_en.html).

UAB made further developments of the INTERPRET DSS independently of RADIONET, partly to facilitate bilateral clinical collaborations. The DMS was modified so that data could be entered on-line, and the data associated with some cases were corrected and re-labeled, which meant that the Java code had to be altered and classifiers retrained (32).

The eTumour Project

eTumour (2004-2009), led by Bernardo Celda at the University of Valencia, was a larger project (ftp://ftp.cordis.europa.eu/pub/lifescihealth/docs/canpr315_en.pdf) that expanded the INTERPRET approach to include multimodal analysis of in vivo MRI and MRSI with ex-vivo transcriptomic (RNA microarrays) and metabolomic (HR-MAS) data obtained from brain tumour biopsies from the same patients in order to explore the potential of all this information, taken together, in predicting response to therapy. Since eTumour involved all the original INTERPRET data-provider partners as well as some new ones (33) it benefited from lessons learned during INTERPRET, particularly with respect to the acquisition protocols, quality control strategies, database, pattern recognition methods and the DSS.

Database preservation

The viDB (http://gabrmn.uab.es/interpretvalidateddb/) can be accessed in two ways. Anyone requesting access for scientific or medical purposes can be granted "view permission", i.e. being able to look at the data but not being able to download either the processed or the unprocessed spectra; it has resulted in about 50 requests. Permission to download the raw data is available upon request to the coordinator, subject to permission from the original data-contributing partners of INTERPRET. This type of access has been granted to five research groups and two consortia (eTumour and HealthAgents (34))and very successful in allowing numerous pattern recognition studies to be performed on the existing data (35-56). The complete iDB has been a relatively unexploited resource, particularly those cases with different degrees of validation and completeness. Both INTERPRET databases are administered and maintained by the UAB team, without specific funding targeting their maintenance.

Using the DMS with the new MRS data formats

The DMS and DMS pipeline have been further developed during the post-INTERPRET period. At UAB, Guillem Mercadal developed the jDMS (32), a Java based MRS format conversion software, which automatically translates any spectrum processed manually with jMRUI into the 512-point INTERPRET canonical format (32). jDMS eliminated the need for constantly updating the MRS format conversion routines within the DMS to cope with the numerous MRS formats and updates that have appeared in the 13 years since the original software was developed, thus keeping the DSS alive and usable. It also permits the use of jMRUI for manual phasing of spectra when the Klose algorithm does not work perfectly. The jDMS was, *de facto*, adopted by the eTumour (33) and HealthAgents (57) multicentre projects, both to display processed spectra, for quality control, and to obtain data matrices for classifier development (35,37-40,42), and also by the CURIAM DSS (58), which derived from these projects. The jDMS also facilitated scientific collaborations with artificial intelligence research groups who could then obtain a consistent, clean data matrix without having to learn how to phase a spectrum or process MRS data.

The DMS pipeline has been recently integrated into the MRUI software as a plug-in named jMRUI2XML (http://gabrmn.uab.es/jmrui2xml), so as to automate the data processing of any SV or MV spectral format, ready for exporting for example in the INTERPRET canonical format into pattern recognition algorithms or the INTERPRET DSS (Figure 4).

Pattern recognition: does the particular method matter?

Sharing the SV INTERPRET data resulted in numerous published studies (35-48,59,60) which mainly developed classifiers or feature extraction methods by using the INTERPRET data either as the training set or as a test set. Both approaches were used in a study performed by eTumour (37), in which 253 pairwise classifiers for GB, MN, ME, and LGG were obtained for

211 short TE spectra from INTERPRET (the training set) and 97 spectra from eight different centres of the eTumour consortium (the test set).

Lessons learned from the pattern-recognition studies

Single voxel

1. Classifiers developed in this way are robust, and perform well on independent data from several centres and manufacturers, acquired at different times by different operators.
2. Most studies addressed relatively easy problems that gave results of around 90% (whether it in terms of accuracy, AUC, or any other measure of performance)(37), no matter what feature extraction or classification methods were employed. The basic paradigm continued to be the "most common tumour types" problem ("LGG vs. AGG vs. MN"), first attempted by (16). Later authors (32,35,40,41,44) developed variations and simplifications of the same problem. In (52), for example, all bilateral combinations of MN, OD, OA, A2, GL, ME, NO, LY, PN, A3 (astrocytomas WHO grade III) and AB (abscesses) were tested. In the main INTERPRET-eTumour paper (37), the seven bilateral combinations were between the GB, ME, MN classes and the AGG superclass. In the latest study using the INTERPRET data, published in 2013 (53), the classes were: "LGG vs. ME", "MN vs. LGG", "LGG vs. GL", "MN vs. GL", "MN vs. NO (normal brain parenchyma)" and "GB vs. ME".
3. Classification boundaries are not necessarily linear when investigating automatic detection of poor quality brain tumour spectra. This was found comparing a Least Squares-Support Vector Machine (LS-SVM) analysis with LDA (61). Also non-linear methods of data reduction may be of help for classification of glial tumour subclasses(60).
4. Difficult problems have more rarely been attempted, the paradigm being "GB vs. ME". Despite having very different (and in the case of ME, very heterogeneous) origins, these two groups of tumours have remarkably similar spectra, so GB and ME were originally

joined into a single group for analysis of multiclass (i.e. more than two classes) problems (24,25). Discrimination of GB from ME is a situation in which the importance of using an independent test set can be observed. Studies that attempted this bilateral discrimination without using a test set, for example (48,52,54), claimed results in the 90% range. In contrast, in most studies that used an independent test set the performance of their classifiers was no better than random (25,37,53). A successful classification was achieved in one study (55) that had three distinctive characteristics: first, training the classifier with INTERPRET data and testing it with an independent test set of 40 cases from three different hospitals and two manufacturers' instruments. Second, the use of a simple and well-known classifier (a single-layer perceptron) but an exhaustive feature selection method. Third was the use of both short and long TE concatenated spectra as input. The best classifier used 4 features of the long TE and one from the short TE spectrum and is available in the latest version of the INTERPRET DSS (http://gabrmn.uab.es/dss).

5. Some outlier cases have been consistently misclassified. Four such outliers (cases I0009, I1390, I0063 and I0450) were first recognized in a study (35) that used an already curated dataset of INTERPRET cases. The same group of outliers was found in the INTERPRET-eTumour study (37). Another study (45) that used Sammon's mapping for visualization of cases, coupled to generative topographic mapping to automatically identify outliers categorized them as artifact-related (outliers caused by artifacts) and class-related (spectra that are outliers with respect to their class).

6. Short and long TE spectra can be concatenated, which can help in some bilateral discriminations, e.g. the "GB vs. ME" problem (55), or for distinguishing meningiomas from other tumour types (35).

7. 3T and 1.5T data, processed with the DMS and converted to the INTERPRET canonical format, can be compatible in terms of classification (39). This study trained a classifier with short TE INTERPRET spectra and tested it with sets of short TE eTumour spectra obtained at 1.5T and 3T. The classification problem was "AGG vs. LGG", either using peak heights or integrated peak areas, and simple classifiers (LDA, KNN and ANN). The results were similar for the 1.5T and the 3T sets, despite the training set consisting of 1.5T spectra.

In a further development of the INTERPRET approach, the UAB group developed SpectraClassifier (62), a program that enables biochemists and other users with no expertise in pattern recognition to make MRS classifiers . It includes simple tools for feature extraction (PCA), selection (greedy stepwise (63)) and classification (Fisher LDA). The system has been successfully used to classify MRSI data of preclinical models (64) and to train classifiers in the last DSS version (32).

Multivoxel

Arjan Simonetti, then at KUN, initiated the post-INTERPRET studies on MV data from the dataset accrued during the project (27); other groups subsequently studied essentially the same group of subjects: 4 volunteers and 20 (28) or 24 patients (in the rest of papers cited later in this section) with MN, OD WHO grade II, III and IV tumours (GB); spectra were acquired at 1.5T (27). In contrast to the SV studies, the main technique for dimensionality reduction was peak area integration for 5 (27), 7 (28), 8 (65) or 10 (66,67) of the main metabolites, although PCA on the whole spectra followed by quadratic discriminant analysis (QDA) was also used (28). The co-registered T1 weighted pre and post gadolinium, T2 weighted and proton density images (28) were also used, either for confirming the result of the MV classification with the anatomy (27), or to improve classification (68-70). The MV data were submitted to a variety of techniques that had previously been used with SV data. Unsupervised studies used clustering

by mixtures of multivariate normal distributions (27), ICA (68,71), Kohonen networks (66), QDA combined with PCA or ICA for feature reduction (28,68), canonical correlation analysis (CCA) (65), support vector machines (SVM) (72), or the latter two combined (67). The following lessons were learned, despite being based on a limited number of patients:

1. A 2D image of the brain could be made showing possible tumour areas and possible tumour heterogeneity in those areas, including some indication of diagnostic reliability.
2. Despite the lower signal-to-noise ratio, classes similar to those handled by the single voxel classifiers could be recognised.
3. The use of co-registered MRI images helped to increase the associated accuracy (69,70).
4. Non-linear methods have better performance in some situations where there is overlap between classes; for example when using a limited set of features from peak integration (10 metabolites) and MRI (4 values from T1, T2, Proton Density and Gadolinium Enhanced T1) to distinguish heterogeneous classes (low grade vs. high grade tumours) (70). The non-linear LS-SVM with a radial basis function also out-performed other methods for more subtle discrimination of glial subclasses in the INTERPRET MV dataset (69).
5. Some of the studies performed recognised the drawback of using supervised methods – particularly to deal with unexpected tumour types. In one study the latter were classified as “unknown”(28).
6. Tumour heterogeneity (73) within a voxel, was tackled in different ways: either by assigning voxels that did not reach a certain threshold of Mahalanobis distance to the centroids of the respective classes as being of “undecided” pathology (28) or by a two-

step process in which tumour typing was followed by a segmentation in which the class “mixed tissue” was introduced (65).

Decision support systems

The SV INTERPRET DSS continued to be developed by UAB (32) and evolved into the current version 3.1, keeping the look and feel of the original. It currently has classifiers for three different problems: “most common tumor types”, “tumour vs. pseudotumoural disease” (74), and “GB vs. ME” (55) and can now be used as jMRUI plugin (Figure 4). It keeps the “make your own overview” feature as well as the possibility of using short, long or short and long TE spectra of the same case. The multiplicity of classifiers now means that the user must have a clear idea of the question to be answered and whether to attempt the analysis with short TE, long TE or combined TE spectra.

Versions 1 and 2 of the DSS have been clinically tested. A retrospective multicentre evaluation of the added value of MRS for diagnosing brain tumours compared the diagnostic accuracy using conventional MRI alone with diagnostic accuracy using the DSS in addition to MRI. Twenty radiologists (only 4 of whom had been concerned with INTERPRET) from 4 European countries, reviewed 16 cases, with access to T1 weighted images prior to and after contrast enhancement, and T2 weighted images before contrast. They then chose one from a list of possible diagnoses and rated its likelihood on a six-point scale, first with the MRI and clinical information and then with additional information from the ^{1}H spectrum analysed and displayed by the DSS. Adding the MRS analysis significantly improved on diagnoses made using MRI alone (AUC_{MRI} =0.88, $AUC_{MRI+MRS}$= 0.92, n=834 readings). AUC values were higher after using the INTERPRET DSS in MN, GB and ME, but only in PN were the AUC significantly different (AUC= 0.50 vs. 0.83 without or with INTERPRET respectively) (16).

Version 1 was also evaluated, together with 3 more systems for spectral classification in a single-centre, prospective study of 40 patients, to see if the MRS information added to MRI analysis improved preoperative diagnosis and grading of brain tumours (75). First, radiologists evaluated MRI and spectroscopists evaluated MRS independently – spectroscopists without any added information related to the patient, radiologists with the usual clinical data. After predicting the tumour type using a 5-point scale, they exchanged predictions and re-evaluated their diagnoses. MRS added value to the preoperative radiological evaluation of the "GB and ME superclass" (AUC_{MRI}= 0.83, $AUC_{MRI+MRS}$= 0.93, n= 12), glial tumours WHO grade III (AUC_{MRI}= 0.70 , $AUC_{MRI+MRS}$= 0.84, n= 12), glial tumours WHO grades II-III (AUC_{MRI}= 0.81, $AUC_{MRI+MRS}$= 0.93, n= 13) and tumours of WHO grade IV (AUC_{MRI}= 0.85, $AUC_{MRI+MRS}$= 0.93 , n= 14). The INTERPRET DSS was the most successful of the four spectral evaluation systems tested, significantly out-performing the radiologists in diagnosis of astrocytomas of WHO grade III (AUC_{MRI}= 0.66 , AUC_{DSS}= 0.87, n= 9 cases).

A lesson learned from (75) was that radiologists in public health system centres do not have the time to embark on formal evaluations of a DSS. Therefore when it was necessary to check whether a new version of the system, with an added long TE classifier, could improve on the prospective study, 6 radiologists with no particular knowledge of spectroscopy and 3 expert MR spectroscopists were left free to use the system to categorise the same 40 cases (76). In most classes and superclasses of brain tumours the AUC were not significantly different from the first study (75) except for astrocytomas WHO grade III, radiologists had an AUC_{DSS} =0.59 (n=54 cases/ 238 readings) whereas the expert spectroscopists reached an AUC_{DSS} =0.71 (n=27 cases /116 readings). This study demonstrated that radiologists with no expertise in spectroscopy can use the system, by adhering to basic guidelines. The ability to choose an echo time other than short TE did not affect results as most classes evaluated did not have significant differences in the AUC reached in comparison to using only one echo time. In

principle, this does not seem to agree with previous literature on classifiers (35,77), however the exact use of the system was not recorded in the latter studies. One important difference is that in the early studies classifiers were developed using only cases from a few common brain tumour types (35,77), whereas in the two more recent evaluations the DSS was used on cases from 15 different WHO classes, some of them as uncommon as melanocytoma or brain lymphoma (75,76).

The DSS version 2 was also evaluated by the SGUL team (78) as a high or low-grade tumour classifier in a short TE (30 ms) single voxel, 1.5 T prospective study on 89 patients with grade IV brain tumors. Prediction of the grade (high/low) with MRI and clinical findings was 82% accurate, whereas with MRS (visual evaluation by a team of expert spectroscopists) and the DSS it was 89% and 84% respectively. All 14 biopsied GB patients were diagnosed as GB by a single neuroradiologist and as high-grade by the DSS, suggesting that the DSS could be used as a confirmatory test in the quite large group of patients for whom brain biopsy would be contraindicated (about 25% in the Barcelona study (75)), such as those with poor functional status or an unfavourable lesion location.

Taken together, these three evaluations show that the system behaves robustly when used in different ways by different types of users, either for visualising the effects of classifiers or for decision-suppport, and importantly, to analyse spectra from pathologies for which there is no specific classifier. In the latter case there will be differences in performance depending on the protocol chosen, which can be critical for certain tumour types, such in as astrocytomas of WHO grade III, and the experience of the person using it.

A recurrent question from users of the DSS has been whether 3T data can be fed into it. Technically, the answer is: yes, using any of the DMS available . However, despite the

encouraging pattern recognition results, lack of 3T data from brain tumours acquired in compatible conditions has hampered that type of clinical evaluation.

As an illustration of the retrospective studies that are possible with the INTERPRET DSS, we used the eTumour database, which contains 7 SV 3T cases with good quality spectra and validated histopathology, all of which have both short and long TE SV data. We processed them with the latest jMRUI plugins mentioned in the previous section (jMRUI2XML and DSS) (Figure 4). Table 1 lists the cases, their diagnoses and the results of using the INTERPRET DSS as a classifier. There was a 100% succcess rate for all classifiers except in the long TE "most common tumour types" classifier where 1 glioblastoma was wrongly classified as LGG (Table 1 and Figure 3~~Figure 3~~). One of the 7 cases was radiation necrosis (et3568), which is not represented as a possible diagnosis, so we considered the classification of AGG to be acceptable in terms of necrosis detection. We could have excluded this case from the analysis, but we decided to keep it since it illustrates the difference between a classifier and a DSS aimed at diagnosis of brain tumours: in a DSS the clinician does not know the diagnosis in advance and therefore may enter cases into the system which belong to other classes than those that the embedded classifiers are able to deal with.

Various DSS versions have been distributed to about 200 users, and the system is currently in clinical use at Uppsala University Hospital (79). Surveys were performed among registered DSS users in 2011 and 2012: 63% of the respondents used the DSS between once a month and once every six months. Half used it to evaluate real cases, the remainder as a learning tool or for teaching; 64% worked in hospitals and 84% in academia (77% worked in academia in collaboration with public clinical centres). All worked with SV spectroscopy and 62% also with MV. A much more extensive survey has been recently performed in the context of the TRANSACT project (http://www.transact-itn.eu/), on the use of jMRUI, the INTERPRET DSS and SpectraClassifier.

ACHIEVEMENTS OF INTERPRET AND POSSIBLE FUTURE DEVELOPMENTS

The INTERPRET collaboration set out to develop a non-invasive method for diagnosing and grading brain tumours using ^{1}H MRS, one that would eventually substitute for the subjective analysis of a biopsy in cases where such biopsy is not available or advisable, and would require no understanding of MRS or biochemistry on the part of the user. It accrued several hundred quality-controlled spectra and associated MRI images, histopathology and clinical details; the uploading and quality control procedures were automated. INTERPRET also developed and tested numerous mathematical classifiers and created a decision support system – the DSS – for presenting the data to the user; finally, the overall prototype system was tested several times. How well did INTERPRET succeed in its original aims?

Successes

A major technical problem solved by the INTERPRET consortium was development of a data input procedure that could cope with the proliferation of incompatible and constantly changing MRS data formats used by the three major manufacturers.

Additional successes were the production of classifier algorithms that worked with raw spectra rather than the metabolite concentrations or integrated peak areas required by previous classification programs, and the demonstration that a useful database could be created by combining spectra acquired on instruments from different manufacturers, using fundamentally different acquisition protocols and numerous software generations. The software classifiers could be “taught” to ignore all these irrelevant issues and focus only on the differences between the tumour classes. It was even possible to classify 3T spectra using a database composed of 1.5T spectra (39).

Several clinical studies have tested the INTERPRET DSS. In view of the limited number of tumour classes that were adequately represented in its database, it was not expected that the prototype would be able to make definitive diagnoses. Despite those initial misgivings, it has performed successfully in several prospective clinical tests, and it is clear that for certain classes of tumour, notably low-grade glial tumours, it significantly improves the diagnostic ability of experienced neuroradiologists.

The need for constantly updating the format conversion routines due to changes in the various manufacturers' data formats was overcome during the project by brute force and later by the jDMS input program. This has allowed the on-going use of the INTERPRET system by many researchers and clinicians. The recent release of the jMRUI2XML jMRUI plugin should solve the problem definitively as long as jMRUI can cope with new spectral formats.

The INTERPRET prototype is still in regular use more than a decade after the project ended. Its database has provided a resource for many studies, particularly to develop new classifiers and it is even used by some clinicians to assist diagnosis. It also has a role in teaching. Furthermore, two subsequent EU collaborative projects, eTumour and HealthAgents were built on the foundations laid by INTERPRET.

Weaknesses

There were a number of obvious weaknesses of the INTERPRET project outcome. Because of the limited size of its database it was necessary to aggregate some tumour types into unfamiliar groupings (e.g. "AGG" or "LGG"), while many rarer tumour types were not represented at all. This obviously limits the routine use of the DSS for clinical diagnosis. The same problem meant that the DSS was not configured to offer a proposed diagnosis with a percentage probability, a feature that some users find desirable, or even essential, and which has subsequently been adopted by eTumour (58). Furthermore, the classifiers developed so

far cannot make some clinically important distinctions, e.g. beween lymphomas and other grade IV brain tumours, or between glial subclasses or grades.

In INTERPRET, 32% of the cases did not have an agreed histopathologic diagnosis (either missing or not agreed) (18). In addition, pathologists are less likely to agree on intermediate grade diagnoses, perhaps because conventional histopathological grading of brain tumours can be difficult (and subjective), particularly for distinguishing intermediate grades (80-83). Preliminary investigation of some of these spectra showed a profile indicative of an intermediate grade. Indeed, some evidence suggests that spectra may be a better indicator of prognosis in terms of survival than the pathology-assigned WHO grade (84,85).

An objection to the whole idea of developing a non-invasive method for brain tumour diagnosis is that "They are all going to be operated on anyway so why not just biopsy them?". However, a significant number of patients with suspicious intracerebral massess are not, or should not be biopsied(78). In addition, efforts are being made to develop more anticancer drugs that will cross the blood-brain barrier. If a solely non-surgical treatment of brain tumours by chemotherapy (and radiotherapy) becomes possible then a non-invasive diagnostic test would have great value.

Can the INTERPRET concept be further developed?

The original impulse to develop INTERPRET came from a desire to find a "killer app" for MRS – that is an application that would be so compelling that it would cause users to purchase the necessary equipment and software. From a health economics perspective, bringing MRS into routine practice requires that the costs associated with the additional imaging and radiologist reporting time are balanced against a significant increase in diagnostic performance and the resulting costs of the downstream investigations and treatments. Although successful in many

ways, INTERPRET did not pass the “killer app” threshold. What would a future INTERPRET need in order to bring MRS into routine use for brain tumour diagnosis?

Clinical needs

Diagnosis prediction: Ideally, the system should diagnose and grade all cancer types and grades that routinely present at a brain cancer centre, although there is no certainty, of course, that all tumour types and grades will have MRS-detectable differential features. The system should also be able to advise a user about which tumour types cannot be distinguished with MRS at each particular magnetic field or set of acquisition conditions. However, to make routine diagnoses , many more classifiers would also have to be developed including: Grade IV tumours (LY vs. PN vs. GB vs. ME) ; Grades of glioma (WHO grades II vs. III vs. IV); Subtypes of glioma, both the classical ones defined by conventional histopathology (OD vs. A2 vs. OA) and the emerging ones based on genetic markers (86,87); Less common tumour types, e.g. hemangiopericytoma (88); Pediatric brain tumours (42,89,90).

A new INTERPRET database should include 3T and eventually 7T spectra, and contain adequate numbers (≥20) of cases of all the major cancer types and grades (although there will always be some cancers so rare that they cannot be scanned in adequate numbers). This database should also include childhood tumours. Given the constant technical improvement in spectroscopic techniques (86,91,92), the database should be an open-ended project, for example to account for data acquired at 7T. There is some evidence of the compatibility of 1.5T MRS classifier being applied to 3T MRS data (39) and that acquisition of additional (well validated) MRS tumour data can be used to further optimise a classifier incrementally without recourse to the full original training data set (40). Hence optimisation could be an ongoing process with more subtle classifiers emerging as enough data of the different tumour types at each field strength is obtained. One solution would be to require users of the system to upload the final diagnosis and clinical details of the case, which would then take its place in the

database. One would have to rely on the local histopathologist's diagnosis and on automated QC procedures.

Prognosis prediction: Ideally the new system should be able to tell whether a tumour type or molecular sub-type is likely to respond to a therapy. Additionally, in glioblastomas, the system should be able to differentiate true progression from pseudoprogression or true response from pseudoresponse (93). Such features would enhance the usefulness of the system, as it would then provide action recommendations, rather than mere assessments (94).

Regional determination of heterogeneity: Another possibility would be to develop a tool for assigning tumour type and grade to each voxel in a spectroscopic image – in effect using the DSS to create a "nosologic image"(28,95,96), either for diagnosis or for prognosis prediction. This is in line with recent data about the invasive assessment of human brain intratumour genetic and epigenetic heterogeneity (97).

Workflow needs

To integrate seamlessly with radiologist workflow and working practices (94) the DSS should fully automate MRS preprocessing, quality control and classification. It should also give percentage likelihoods for each prediction and give diagnoses in terms that are familiar to clinicians – e.g."oligodendroglioma", or "low grade glial tumour" rather than "LGG". The system could also tutor users in making use of and understanding the significance of MRS data and motivate them through game-like diagnostic challenges that would demonstrate the added value of MRS data (98). But such a system must fully integrate with the hospitals' picture archiving and commnication sytem (PACS) used for standard radiological images, so that any report and data visualisation based on the DSS output can be stored along with the routine MRI data. It is feasible for the MRS(I) and relevant MRI data to be automatically re-directed to a standalone system for analysis and automated generation of nosologic images, for example, that are then returned to the system for viewing and archiving with the standard

MRI in almost real-time. Alternatively, the DSS software could provide post-acquisition analysis, with previously archived MRS(I) and MRI data taken from PACS for processing, analysis and generation of an expert report to be subsequently archived to PACS.

Secondly, who would own and maintain this new DSS and its large database? One solution would be to develop a commercial DSS that could be sold to users. The ownership of the database necessary for the use of that DSS is a more difficult question. The current iDB and viDB consist of de-identified spectra obtained from patients in many different countries. All these patients gave informed consent, but numerous national legal frameworks were involved. These databases are owned by the original INTERPRET partners who allow them to be used by the INTERPRET DSS. If a larger database is to be developed it would probably be simplest if it is also owned and maintained by a non-commercial entity, as it would be challenging (and probably expensive) to develop a legal framework that allowed its commercial ownership. One possible solution would be to integrate it into a large data infrastructure such as Elixir, which is currently under construction (http://www.elixir-europe.org/about/rationale). An alternative approach to these technological and medico-legal challenges, would be not to have a centralised database but rather a cloud database, whereby each contributing centre owns and manages its own local database, sharing only de-identified spectral features that are relevant for classification. This alternative approach poses the technological challenge of redesigning the classification algorithms to use distributed datasets, which is already an active area of research (99,100). This concept, using software agents (34) had already been explored during HealthAgents (41,101,102) but, unfortunately, with no practical continuity or commercial exploitation after the funding period ended.

All the studies so far have been designed to demonstrate that use of MRS data by the prototype DSS improves diagnostic power. A DSS for practical clinical use should have

classifiers that use both MRS data, clinical data and MRI data. Simple parameters such as the patient's age and sex are already available in the database and could significantly refine certain diagnoses. Another source of information that could be exploited in a future system would be the MRI images (or indeed any other images) of the tumour. Texture analysis software or morphological analysis could produce quantitative parameters that could be used by a classifier (103-105).

One final speculation: it has always been tacitly assumed that the aim of INTERPRET and its subsequent developments was to reproduce the diagnosis made by a histopathologist. In reality, it is not impossible that MRS-based classification of cancers might *improve* on histopathology. MRS detects changes in the spectral pattern caused by small-molecule metabolites such as 2-hydroxyglutarate, lipids and even macromolecules in the living tumour, all of which contain a different type of information from that available on the classical microscope slides used for histopathology, or even from the genomic data that are nowadays coming into use for tumour classification (87,106-109).

Perhaps, for instance, MRS-based classifiers could be devised that will provide a more accurate prognosis for some brain tumour sub-types (84,110,111), or could define sub-categories that would help to personalize the treatment of patients. Studies of that type will require another level of information gathering. Long-term data on disease progression and treatment response will have to be collected for the patients whose spectra are in the database and then classifiers constructed that will act as prognostic (i.e. course of disease) and predictive (i.e. response to treatment) biomarkers. Finally, it will be necessary to make prospective predictions by using those classifiers on a new cohort of patients and then to wait several years and see whether the predictions were accurate. The big challenge here would be to develop strategies that may work, at least partially, in a non-supervised or semi-supervised way (50).For any of these strategies to have a direct translation into clinical practice, it is

mandatory that studies are performed according to evidence based medicine (EBM) criteria such as STARD (112), in contrast to most clinical MRS literature (reviewed by (113)).Otherwise, MRS would risk denial of reimbursement as happened some years ago (CPT 76390) across hundreds of healthcare providers in the USA (113). For this, efforts in standardising MRS data acquisition and analysis, and in particular the implementation of technical improvements to enable good quality MRS acquisitions to be performed faster and robustly, as well as clinical guidelines for MRS use should be encouraged (114).

MRS is already more than 30 years old: let us hope that its future is long enough for these possibilities to be realized.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Drs. Marinette van der Graaf, Chantal Remy, Arend Heerschap and Des Watson for their comments which helped to improve this manuscript.

MJS and CA acknowledge project MARESCAN (SAF2011-23870) and MOLIMAGLIO (SAF2014-52332-R) from Ministerio de Economia y Competitividad in Spain. This work was also partially funded by CIBER-BBN, which is an initiative of the VI National R&D&i Plan *2008-2011, CIBER Actions* and financed by the *Instituto de Salud Carlos III* with assistance from the European Regional Development Fund. JRG acknowledges support from Cancer Research UK, grant C1459/A2592, the University of Cambridge and Hutchison Whampoa Ltd.

REFERENCES

1. Griffiths JR, Cady E, Edwards RH, McCready VR, Wilkie DR, Wiltshaw E. 31P-NMR studies of a human tumour in situ. Lancet 1983;1(8339):1435-1436.
2. Howells SL, Maxwell RJ, Peet AC, Griffiths JR. An investigation of tumor 1H nuclear magnetic resonance spectra by the application of chemometric techniques. Magn Reson Med 1992;28(2):214-236.

3. Howells SL, Maxwell RJ, Howe FA, Peet AC, Stubbs M, Rodrigues LM, Robinson SP, Baluch S, Griffiths JR. Pattern recognition of 31P magnetic resonance spectroscopy tumour spectra obtained in vivo. NMR Biomed 1993;6(4):237-241.
4. Preul MC, Caramanos Z, Collins DL, Villemure JG, Leblanc R, Olivier A, Pokrupa R, Arnold DL. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. Nat Med 1996;2(3):323-325.
5. Usenius JP, Tuohimetsa S, Vainio P, Ala-Korpela M, Hiltunen Y, Kauppinen RA. Automated classification of human brain tumours by neural network analysis using in vivo 1H magnetic resonance spectroscopic metabolite phenotypes. Neuroreport 1996;7(10):1597-1600.
6. Howells SL, Maxwell RJ, Griffiths JR. Classification of tumour 1H NMR spectra by pattern recognition. NMR Biomed 1992;5(2):59-64.
7. Tate AR, Griffiths JR, Martinez-Perez I, Moreno A, Barba I, Cabanas ME, Watson D, Alonso J, Bartumeus F, Isamat F, Ferrer I, Vila F, Ferrer E, Capdevila A, Arus C. Towards a method for automated classification of 1H MRS spectra from brain tumours. NMR Biomed 1998;11(4-5):177-191.
8. Sharples M, Jeffery NP, du Boulay B, Teather BA, Teather D, du Boulay GH. Structured computer-based training in the interpretation of neuroradiological images. Int J Med Inform 2000;60(3):263-280.
9. Favre J, Taha JM, Burchiel KJ. An analysis of the respective risks of hematoma formation in 361 consecutive morphological and functional stereotactic procedures. Neurosurgery 2002;50(1):48-56; discussion 56-47.
10. Hall WA. The safety and efficacy of stereotactic biopsy for intracranial lesions. Cancer 1998;82(9):1749-1755.
11. Field M, Witham TF, Flickinger JC, Kondziolka D, Lunsford LD. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. J Neurosurg 2001;94(4):545-551.
12. Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arus C. Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. J Neurosurg 2006;105(1):6-14.
13. Witjes H, Melssen WJ, in 't Zandt HJ, van der Graaf M, Heerschap A, Buydens LM. Automatic correction for phase shifts, frequency shifts, and lineshape distortions across a series of single resonance lines in large spectral data sets. J Magn Reson 2000;144(1):35-44.
14. Simonetti AW, Melssen WJ, van der Graaf M, Heerschap A, Buydens LM. Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. J Magn Reson 2002;159(2):151-157.
15. van der Graaf M, Julia-Sape M, Howe FA, Ziegler A, Majos C, Moreno-Torres A, Rijpkema M, Acosta D, Opstad KS, van der Meulen YM, Arus C, Heerschap A. MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. NMR Biomed 2008;21(2):148-158.
16. Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arus C. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR Biomed 2006;19(4):411-434.
17. Murphy M, Loosemore A, Ferrer I, Wesseling P, Wilkins PR, Bell BA. Neuropathological diagnostic accuracy. Br J Neurosurg 2002;16(5):461-464.

18. Julia-Sape M, Acosta D, Mier M, Arus C, Watson D. A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. Magn Reson Mater Phy 2006;19(1):22-33.
19. Tate AR, Majos C, Moreno A, Howe FA, Griffiths JR, Arus C. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. Magn Reson Med 2003;49(1):29-36.
20. Altman DG, Royston P. What do we mean by validating a prognostic model? Statistics in Medicine 2000;19(4):453-473.
21. Broadhurst D, Kell D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006;2(4):171-196.
22. Ladroue C, Howe FA, Griffiths JR, Tate AR. Independent component analysis for automated decomposition of in vivo magnetic resonance spectra. Magn Reson Med 2003;50(4):697-703.
23. Ladroue CLC. Pattern recognition techniques for the study of magnetic resonance spectra of brain tumours. London: University of London; 2004.
24. Devos A, Lukas L, Suykens JA, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arus C, Van Huffel S. Classification of brain tumours using short echo time 1H MR spectra. J Magn Reson 2004;170(1):164-175.
25. Lukas L, Devos A, Suykens JA, Vanhamme L, Howe FA, Majos C, Moreno-Torres A, Van der Graaf M, Tate AR, Arus C, Van Huffel S. Brain tumor classification based on long echo proton MRS signals. Artif Intell Med 2004;31(1):73-89.
26. Witjes H, Rijpkema M, van der Graaf M, Melssen W, Heerschap A, Buydens L. Multispectral magnetic resonance image analysis using principal component and linear discriminant analysis. J Magn Reson Imaging 2003;17(2):261-269.
27. Wehrens R, Simonetti AW, Buydens LMC. Mixture modelling of medical magnetic resonance data. Journal of Chemometrics 2002;16(6):274-282.
28. Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LM. A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy. Anal Chem 2003;75(20):5352-5361.
29. Simonetti A. Investigation of brain tumor classification and its reliability using chemometrics on MR spectroscopy and MR imaging data. Nijmegen, the Netherlands: Radboud University Nijmegen; 2004.
30. Underwood J, Luckin R, Cox R, Watson D, Tate R. Focussing User Studies: Requirements Capture for a Decision Support Tool. 2000. Citeseer. p 88-92.
31. Underwood J, Tate AR, Luckin R, Majos C, Capdevila A, Howe F, Griffiths J, Arus C. A prototype decision support system for MR spectroscopy-assisted diagnosis of brain tumours. Stud Health Technol Inform 2001;84(Pt 1):561-565.
32. Perez-Ruiz A, Julia-Sape M, Mercadal G, Olier I, Majos C, Arus C. The INTERPRET Decision-Support System version 3.0 for evaluation of Magnetic Resonance Spectroscopy data from human brain tumours and other abnormal brain masses. BMC Bioinformatics 2010;11(1):581.
33. Julià-Sapé M, Lurgi M, Mier M, Estanyol F, Rafael X, Candiota A, Barceló A, García A, Martínez-Bisbal M, Ferrer-Luna R, Moreno-Torres Á, Celda B, Arús C. Strategies for annotation and curation of translational databases: the eTUMOUR project. Database (Oxford) 2012;2012:bas035.
34. González-Vélez H, Mier M, Julià-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van Huffel S, Lluch-Ariet M. HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. Applied Intelligence 2009;30(3):191-202.

35. Garcia-Gomez JM, Tortajada S, Vidal C, Julia-Sape M, Luts J, Moreno-Torres A, Van Huffel S, Arus C, Robles M. The effect of combining two echo times in automatic brain tumor classification by MRS. NMR Biomed 2008;21(10):1112-1125.
36. Luts J, Poullet JB, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JA, Van Huffel S. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. Magn Reson Med 2008;60(2):288-298.
37. Garcia-Gomez JM, Luts J, Julia-Sape M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-Garcia E, Olier I, Postma G, Monleon D, Moreno-Torres A, Pujol J, Candiota AP, Martinez-Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arus C, Robles M. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. Magn Reson Mater Phy 2009;22(1):5-18.
38. Vicente J, García-Gómez J, Tortajada S, Navarro A, Howe F, Peet A, Julià-Sapé M, Celda B, Wesseling P, Lluch-Ariet M, Robles M. Ranking of Brain Tumour Classifiers Using a Bayesian Approach. In: Cabestany J, Sandoval F, Prieto A, Corchado J, editors. Bio-Inspired Systems: Computational and Ambient Intelligence. Volume 5517, Lecture Notes in Computer Science: Springer Berlin / Heidelberg; 2009. p 1005-1012.
39. Fuster-Garcia E, Navarro C, Vicente J, Tortajada S, García-Gómez J, Sáez C, Calvar J, Griffiths J, Julià-Sapé M, Howe F, Pujol J, Peet A, Heerschap A, Moreno-Torres À, Martínez-Bisbal MC, Martínez-Granados B, Wesseling P, Semmler W, Capellades J, Majós C, Alberich-Bayarri À, Capdevila A, Monleón D, Martí-Bonmatí L, Arús C, Celda B, Robles M. Compatibility between 3T 1H SV-MRS data and automatic brain tumour diagnosis support systems based on databases of 1.5T 1H SV-MRS spectra. Magn Reson Mater Phy 2011;24(1):35-42.
40. Tortajada S, Fuster-Garcia E, Vicente J, Wesseling P, Howe FA, Julia-Sape M, Candiota AP, Monleon D, Moreno-Torres A, Pujol J, Griffiths JR, Wright A, Peet AC, Martinez-Bisbal MC, Celda B, Arus C, Robles M, Garcia-Gomez JM. Incremental Gaussian Discriminant Analysis based on Graybill and Deal weighted combination of estimators for brain tumour diagnosis. J Biomed Inform 2011;44(4):677-687.
41. Saez C, Garcia-Gomez JM, Vicente J, Tortajada S, Luts J, Dupplaw D, Van Huffel S, Robles M. A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents. The Knowledge Engineering Review 2011;26(Special Issue 03):283-301.
42. Vicente J, Fuster-Garcia E, Tortajada S, Garcia-Gomez JM, Davies N, Natarajan K, Wilson M, Grundy RG, Wesseling P, Monleon D, Celda B, Robles M, Peet AC. Accurate classification of childhood brain tumours by in vivo (1)H MRS - a multi-centre study. Eur J Cancer 2013;49(3):658-667.
43. Vellido A, Julià-Sapé M, Romero E, Arús C. Exploratory Characterization of Outliers in a Multi-centre 1H-MRS Brain Tumour Dataset. In: Lovrek I, Howlett R, Jain L, editors. Knowledge-Based Intelligent Information and Engineering Systems. Volume 5178, Lecture Notes in Computer Science: Springer Berlin / Heidelberg; 2008. p 189-196.
44. Nebot À, Castro F, Vellido A, Julià-Sapé M, Arús C. Rule-Based Assistance to Brain Tumour Diagnosis Using LR-FIR. In: Lovrek I, Howlett R, Jain L, editors. Knowledge-Based Intelligent Information and Engineering Systems. Volume 5178, Lecture Notes in Computer Science: Springer Berlin / Heidelberg; 2008. p 173-180.
45. Vellido A, Romero E, González-Navarro FF, Belanche-Muñoz LA, Julià-Sapé M, Arús C. Outlier exploration and diagnostic classification of a multi-centre 1H-MRS brain tumour database. Neurocomputing 2009;72(13-15):3085-3097.
46. González-Navarro FF, Belanche-Muñoz LA, Romero E, Vellido A, Julià-Sapé M, Arús C. Feature and model selection with discriminatory visualization for diagnostic classification of brain tumors. Neurocomputing 2010;73(4-6):622-632.

47. Cruz-Barbosa R, Vellido A. Semi-supervised analysis of human brain tumours from partially labeled MRS information, using manifold learning models. Int J Neural Syst 2010;21(1):17-29.
48. Colas F, Kok JN, Vellido A. Finding discriminative subtypes of aggressive brain tumours using magnetic resonance spectroscopy. Conf Proc IEEE Eng Med Biol Soc 2010;2010:1065-1068.
49. Arizmendi C, Sierra DA, Vellido A, Romero E. Automated classification of brain tumours from short echo time in vivo MRS data using Gaussian Decomposition and Bayesian Neural Networks. Expert Systems with Applications 2014;41(11):5296-5307.
50. Ortega-Martorell S, Ruiz H, Vellido A, Olier I, Romero E, Julia-Sape M, Martin JD, Jarman IH, Arus C, Lisboa PJ. A Novel Semi-Supervised Methodology for Extracting Tumor Type-Specific MRS Sources in Human Brain Data. PLoS One 2013;8(12):e83773.
51. Vilamala A, Lisboa PJG, Ortega-Martorell S, Vellido A. Discriminant Convex Non-negative Matrix Factorization for the classification of human brain tumours. Pattern Recognition Letters 2013;34(14):1734-1747.
52. Arizmendi C, Hernandez-Tamames J, Romero E, Vellido A, Del Pozo F. Diagnosis of brain tumours from magnetic resonance spectroscopy using wavelets and Neural Networks. Conf Proc IEEE Eng Med Biol Soc 2010;2010:6074-6077.
53. Fuster-Garcia E, Tortajada S, Vicente J, Robles M, Garcia-Gomez JM. Extracting MRS discriminant functional features of brain tumors. NMR Biomed 2013;26(5):578-592.
54. Arizmendi C, Sierra DA, Vellido A, Romero E. Brain tumour classification using Gaussian decomposition and neural networks. Conf Proc IEEE Eng Med Biol Soc 2012;2011:5645-5648.
55. Vellido A, Romero E, Julia-Sape M, Majos C, Moreno-Torres A, Pujol J, Arus C. Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel (1)H MRS. NMR Biomed 2012;25(6):819-828.
56. Berger M, Sembritzki K, Hornegger J, Bauer C. Increasing the credibility of MR spectroscopy-based automatic brain tumor classification systems. Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on; 2014. p 345-348.
57. Estanyol F, Rafael X, Roset R, Lurgi M, Mier M, Lluch-Ariet M. A Web-accessible distributed data warehouse for brain tumour diagnosis. The Knowledge Engineering Review 2011;26(Special Issue 03):329-351.
58. Sáez C, Martí-Bonmatí L, Alberich-Bayarri Á, Robles M, García-Gómez JM. Randomized pilot study and qualitative evaluation of a clinical decision support system for brain tumour diagnosis based on SV 1H MRS: Evaluation as an additional information procedure for novice radiologists. Computers in Biology and Medicine 2014;45(0):26-33.
59. Yang G, Nawaz T, Barrick T, Howe F, Slabaugh G. Discrete Wavelet Transform Based Whole-Spectral and Sub-Spectral Analysis for Improved Brain Tumour Clustering using Single Voxel MR Spectroscopy. Biomedical Engineering, IEEE Transactions on 2015;PP(99):1-1.
60. Yang G, Raschke F, Barrick TR, Howe FA. Manifold Learning in MR spectroscopy using nonlinear dimensionality reduction and unsupervised clustering. Magnetic Resonance in Medicine 2014:n/a-n/a.
61. Wright AJ, Arus C, Wijnen JP, Moreno-Torres A, Griffiths JR, Celda B, Howe FA. Automated quality control protocol for MR spectra of brain tumors. Magn Reson Med 2008;59(6):1274-1281.
62. Ortega-Martorell S, Olier I, Julia-Sape M, Arus C. SpectraClassifier 1.0: a user friendly, automated MRS-based classifier-development system. BMC Bioinformatics 2010;11.

63. Hall MA. Correlation-based feature subset selection for machine learning. Hamilton, New Zealand: Waikato; 1999. 198 p.
64. Simoes RV, Ortega-Martorell S, Delgado-Goni T, Le Fur Y, Pumarola M, Candiota AP, Martin J, Stoyanova R, Cozzone PJ, Julia-Sape M, Arus C. Improving the classification of brain tumors in mice with perturbation enhanced (PE)-MRSI. Integr Biol (Camb) 2012;4(2):183-191.
65. De Vos M, Laudadio T, Simonetti AW, Heerschap A, Van Huffel S. Fast nosologic imaging of the brain. Journal of Magnetic Resonance 2007;184(2):292-301.
66. Melssen W, Wehrens R, Buydens L. Supervised Kohonen networks for classification problems. Chemometrics and Intelligent Laboratory Systems 2006;83(2):99-113.
67. Luts J, Laudadio T, Idema AJ, Simonetti AW, Heerschap A, Vandermeulen D, Suykens JAK, Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. NMR in Biomedicine 2009;22(4):374-390.
68. Simonetti AW, Melssen WJ, Szabo de Edelenyi F, van Asten JJ, Heerschap A, Buydens LM. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. NMR Biomed 2005;18(1):34-43.
69. Luts J, Heerschap A, Suykens JA, Van Huffel S. A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. Artif Intell Med 2007;40(2):87-102.
70. Devos A, Simonetti AW, van der Graaf M, Lukas L, Suykens JA, Vanhamme L, Buydens LM, Heerschap A, Van Huffel S. The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. J Magn Reson 2005;173(2):218-228.
71. Szabo de Edelenyi F, Simonetti AW, Postma G, Huo R, Buydens LMC. Application of independent component analysis to 1H MR spectroscopic imaging exams of brain tumours. Analytica Chimica Acta 2005;544(1–2):36-46.
72. Krooshof PWT, Üstün Bl, Postma GJ, Buydens LMC. Visualization and Recovery of the (Bio)chemical Interesting Variables in Data Analysis with Support Vector Machine Classification. Analytical Chemistry 2010;82(16):7000-7007.
73. Segebarth CM, Balériaux DF, Luyten PR, Den Hollander JA. Detection of metabolic heterogeneity of human intracranial tumors in vivo by 1h nmr spectroscopic imaging. Magnetic Resonance in Medicine 1990;13(1):62-76.
74. Majos C, Aguilera C, Alonso J, Julia-Sape M, Castaner S, Sanchez JJ, Samitier A, Leon A, Rovira A, Arus C. Proton MR spectroscopy improves discrimination between tumor and pseudotumoral lesion in solid brain masses. AJNR Am J Neuroradiol 2009;30(3):544-551.
75. Julia-Sape M, Coronel I, Majos C, Candiota AP, Serrallonga M, Cos M, Aguilera C, Acebes JJ, Griffiths JR, Arus C. Prospective diagnostic performance evaluation of single-voxel 1H MRS for typing and grading of brain tumours. NMR Biomed 2012;25(4):661-673.
76. Julia-Sape M, Majos C, Camins A, Samitier A, Baquero M, Serrallonga M, Domenech S, Grive E, Howe FA, Opstad K, Calvar J, Aguilera C, Arus C. Multicentre evaluation of the INTERPRET decision support system 2.0 for brain tumour classification. NMR Biomed 2014;27(9):1009-1018.
77. Majos C, Julia-Sape M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arus C, Gili J. Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. AJNR Am J Neuroradiol 2004;25(10):1696-1704.
78. Fellows GA, Wright AJ, Sibtain NA, Rich P, Opstad KS, McIntyre DJ, Bell BA, Griffiths JR, Howe FA. Combined use of neuroradiology and 1H-MR spectroscopy may provide an

intervention limiting diagnosis of glioblastoma multiforme. J Magn Reson Imaging 2010;32(5):1038-1044.
79. Weis J, Ring P, Olofsson T, Ortiz-Nieto F, Wikström J. Short echo time MR spectroscopy of brain tumors: Grading of cerebral gliomas by correlation analysis of normalized spectral amplitudes. Journal of Magnetic Resonance Imaging 2010;31(1):39-45.
80. Coons SW, Johnson PC. Regional heterogeneity in the proliferative activity of human gliomas as measured by the Ki-67 labeling index. J Neuropathol Exp Neurol 1993;52(6):609-618.
81. Jackson RJ, Fuller GN, Abi-Said D, Lang FF, Gokaslan ZL, Shi WM, Wildrick DM, Sawaya R. Limitations of stereotactic biopsy in the initial management of gliomas. Neuro Oncol 2001;3(3):193-200.
82. Prayson RA, Agamanolis DP, Cohen ML, Estes ML, Kleinschmidt-DeMasters BK, Abdul-Karim F, McClure SP, Sebek BA, Vinay R. Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading. J Neurol Sci 2000;175(1):33-39.
83. Kros JM, Gorlia T, Kouwenhoven MC, Zheng PP, Collins VP, Figarella-Branger D, Giangaspero F, Giannini C, Mokhtari K, Mork SJ, Paetau A, Reifenberger G, van den Bent MJ. Panel review of anaplastic oligodendroglioma from European Organization For Research and Treatment of Cancer Trial 26951: assessment of consensus in diagnosis, influence of 1p/19q loss, and correlations with outcome. J Neuropathol Exp Neurol 2007;66(6):545-551.
84. Majos C, Bruna J, Julia-Sape M, Cos M, Camins A, Gil M, Acebes JJ, Aguilera C, Arus C. Proton MR spectroscopy provides relevant prognostic information in high-grade astrocytomas. AJNR Am J Neuroradiol 2011;32(1):74-80.
85. Herminghaus S, Dierks T, Pilatus U, Moller-Hartmann W, Wittsack J, Marquardt G, Labisch C, Lanfermann H, Schlote W, Zanella FE. Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data. J Neurosurg 2003;98(1):74-81.
86. Andronesi OC, Kim GS, Gerstner E, Batchelor T, Tzika AA, Fantin VR, Vander Heiden MG, Sorensen AG. Detection of 2-hydroxyglutarate in IDH-mutated glioma patients by in vivo spectral-editing and 2D correlation magnetic resonance spectroscopy. Sci Transl Med 2012;4(116):116ra114.
87. Andronesi OC, Rapalino O, Gerstner E, Chi A, Batchelor TT, Cahill DP, Sorensen AG, Rosen BR. Detection of oncogenic IDH1 mutations using magnetic resonance spectroscopy of 2-hydroxyglutarate. The Journal of Clinical Investigation 2013;123(9):3659-3663.
88. Barba I, Moreno A, Martinez-Perez I, Tate AR, Cabanas ME, Baquero M, Capdevila A, Arus C. Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas. J Neurosurg 2001;94(1):55-60.
89. Orphanidou-Vlachou E, Auer D, Brundler MA, Davies NP, Jaspan T, MacPherson L, Natarajan K, Sun Y, Arvanitis TN, Grundy RG, Peet AC. (1)H magnetic resonance spectroscopy in the diagnosis of paediatric low grade brain tumours. European journal of radiology 2013;82(6):e295-301.
90. Gill SK, Wilson M, Davies NP, Macpherson L, English M, Arvanitis TN, Peet AC. Diagnosing relapse in children's brain tumors using metabolite profiles. Neuro Oncol 2014;16(1):156-164.
91. Nelson SJ, Kurhanewicz J, Vigneron DB, Larson PEZ, Harzstark AL, Ferrone M, van Criekinge M, Chang JW, Bok R, Park I, Reed G, Carvajal L, Small EJ, Munster P, Weinberg VK, Ardenkjaer-Larsen JH, Chen AP, Hurd RE, Odegardstuen L-I, Robb FJ,

Tropp J, Murray JA. Metabolic Imaging of Patients with Prostate Cancer Using Hyperpolarized [1-13C]Pyruvate. Science Translational Medicine 2013;5(198):198ra108.
92. Scheenen TJ, Heerschap A, Klomp DJ. Towards 1H-MRSI of the human brain at 7T with slice-selective adiabatic refocusing pulses. Magn Reson Mater Phy 2008;21(1-2):95-101.
93. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, DeGroot J, Wick W, Gilbert MR, Lassman AB, Tsien C, Mikkelsen T, Wong ET, Chamberlain MC, Stupp R, Lamborn KR, Vogelbaum MA, van den Bent MJ, Chang SM. Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group. Journal of Clinical Oncology 2010;28(11):1963-1972.
94. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005;330(7494):765.
95. De Edelenyi FS, Rubin C, Esteve F, Grand S, Decorps M, Lefournier V, Le Bas JF, Remy C. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. Nat Med 2000;6(11):1287-1289.
96. Luts J, Laudadio T, Idema AJ, Simonetti AW, Heerschap A, Vandermeulen D, Suykens JA, Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. NMR Biomed 2009;22(4):374-390.
97. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, Curtis C, Watts C, Tavare S. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A 2013;110(10):4009-4014.
98. Underwood J, Luckin R. Adequate Decision Support Systems Must Also Be Good Learning Environments. In: Cerri S, Gouardères G, Paraguaçu F, editors. Intelligent Tutoring Systems. Volume 2363, Lecture Notes in Computer Science: Springer Berlin Heidelberg; 2002. p 1012-1012.
99. Kargupta H. Next generation of data mining. Boca Raton: CRC Press; 2009. xxiv, 605 p., 608 p. of plates p.
100. Samanthula BK, Elmehdwi Y, Jiang W. k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data. CoRR. Volume abs/1403.5001. http://arxiv.org/abs/1403.5001; 2014.
101. Xiao L, Dasmahapatra S, Lewis P, Hu B, Peet A, Gibb A, Dupplaw D, Croitoru M, Estanyol F, Martinez-Miranda J, Gonzalez-Velez H, Lluch-Ariet M. The design and implementation of a novel security model for HealthAgents. The Knowledge Engineering Review 2011;26(Special Issue 03):261-282.
102. Hu B, Croitoru M, Roset R, Dupplaw D, Lurgi M, Dasmahapatra S, Lewis P, Martínez-Miranda J, SÃ¡ez C. The HealthAgents ontology: knowledge representation in a distributed decision support system for brain tumours. The Knowledge Engineering Review 2011;26(Special Issue 03):303-328.
103. Herlidou-Meme S, Constans JM, Carsin B, Olivie D, Eliat PA, Nadal-Desbarats L, Gondry C, Le Rumeur E, Idy-Peretti I, de Certaines JD. MRI texture analysis on texture test objects, normal brain and intracranial tumors. Magn Reson Imaging 2003;21(9):989-993.
104. Rodriguez Gutierrez D, Awwad A, Meijer L, Manita M, Jaspan T, Dineen RA, Grundy RG, Auer DP. Metrics and textural features of MRI diffusion to improve classification of pediatric posterior fossa tumors. AJNR Am J Neuroradiol 2014;35(5):1009-1015.
105. Blanchet L, Krooshof PWT, Postma GJ, Idema AJ, Goraj B, Heerschap A, Buydens LMC. Discrimination between Metastasis and Glioblastoma Multiforme Based on

Morphometric Analysis of MR Images. American Journal of Neuroradiology 2011;32(1):67-73.
106. Brandes AA, Franceschi E, Tosoni A, Blatt V, Pession A, Tallini G, Bertorelle R, Bartolini S, Calbucci F, Andreoli A, Frezza G, Leonardi M, Spagnolli F, Ermani M. MGMT promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. J Clin Oncol 2008;26(13):2192-2197.
107. Reifenberger G, Weber RG, Riehmer V, Kaulich K, Willscher E, Wirth H, Gietzelt J, Hentschel B, Westphal M, Simon M, Schackert G, Schramm J, Matschke J, Sabel MC, Gramatzki D, Felsberg J, Hartmann C, Steinbach JP, Schlegel U, Wick W, Radlwimmer B, Pietsch T, Tonn JC, von Deimling A, Binder H, Weller M, Loeffler M, German Glioma N. Molecular characterization of long-term survivors of glioblastoma using genome- and transcriptome-wide profiling. International journal of cancer Journal international du cancer 2014;135(8):1822-1831.
108. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS, Thompson CB, Kaufman A, Guryanova O, Levine R, Heguy A, Viale A, Morris LG, Huse JT, I.K. M, Chan TA. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 2012;483:479-483.
109. Brennan Cameron W, Verhaak Roel GW, McKenna A, Campos B, Noushmehr H, Salama Sofie R, Zheng S, Chakravarty D, Sanborn JZ, Berman Samuel H, Beroukhim R, Bernard B, Wu C-J, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla Sachet A, Ciriello G, Yung WK, Zhang W, Sougnez C, Mikkelsen T, Aldape K, Bigner Darell D, Van Meir Erwin G, Prados M, Sloan A, Black Keith L, Eschbacher J, Finocchiaro G, Friedman W, Andrews David W, Guha A, Iacocca M, O'Neill Brian P, Foltz G, Myers J, Weisenberger Daniel J, Penny R, Kucherlapati R, Perou Charles M, Hayes DN, Gibbs R, Marra M, Mills Gordon B, Lander E, Spellman P, Wilson R, Sander C, Weinstein J, Meyerson M, Gabriel S, Laird Peter W, Haussler D, Getz G, Chin L. The Somatic Genomic Landscape of Glioblastoma. Cell;157(3):753.
110. Hattingen E, Raab P, Franz K, Lanfermann H, Setzer M, Gerlach R, Zanella F, Pilatus U. Prognostic value of choline and creatine in WHO grade II gliomas. Neuroradiology 2008;50(9):759-767.
111. Pope WB, Prins RM, Albert Thomas M, Nagarajan R, Yen KE, Bittinger MA, Salamon N, Chou AP, Yong WH, Soto H, Wilson N, Driggers E, Jang HG, Su SM, Schenkein DP, Lai A, Cloughesy TF, Kornblum HI, Wu H, Fantin VR, Liau LM. Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy. J Neurooncol 2011;107(1):197-205.
112. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology 2003;226(1):24-28.
113. Lin AP, Tran TT, Ross BD. Impact of evidence-based medicine on magnetic resonance spectroscopy. NMR Biomed 2006;19(4):476-483.
114. Öz G, Alger JR, Barker PB, Bartha R, Bizzi A, Boesch C, Bolan PJ, Brindle KM, Cudalbu C, Dinçer A, Dydak U, Emir UE, Frahm J, González RG, Gruber S, Gruetter R, Gupta RK, Heerschap A, Henning A, Hetherington HP, Howe FA, Hüppi PS, Hurd RE, Kantarci K, Klomp DWJ, Kreis R, Kruiskamp MJ, Leach MO, Lin AP, Luijten PR, Marjańska M, Maudsley AA, Meyerhoff DJ, Mountford CE, Nelson SJ, Pamir MN, Pan JW, Peet AC, Poptani H, Posse S, Pouwels PJW, Ratai E-M, Ross BD, Scheenen TWJ, Schuster C, Smith ICP, Soher BJ, Tkáč I, Vigneron DB, Kauppinen RA. Clinical Proton MR Spectroscopy in Central Nervous System Disorders. Radiology 2014;270(3):658-679.

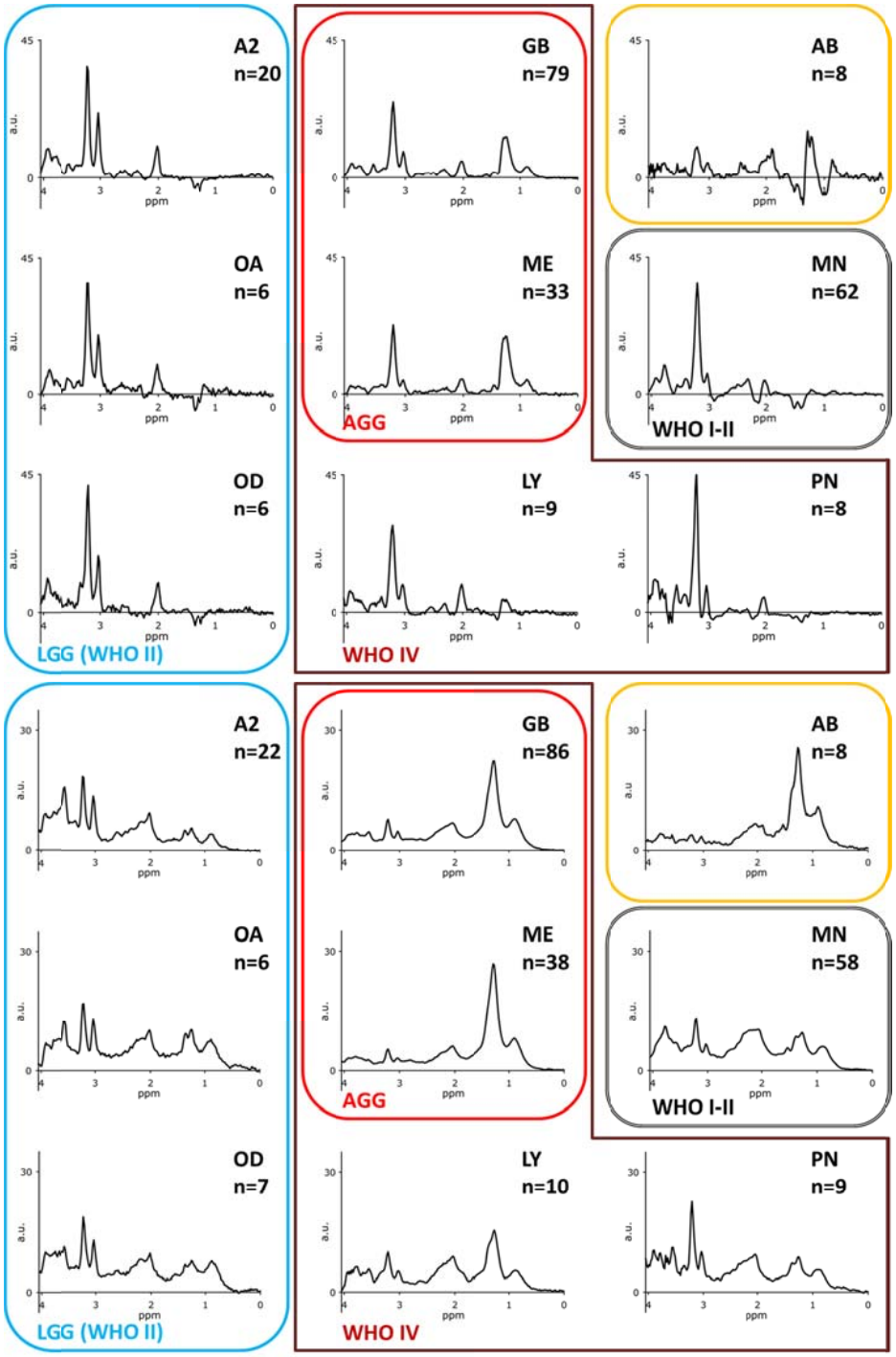

Figure 1. Top, mean (i.e. average) short TE (20-32 msec) spectra; bottom, mean long TE (135-144 msec) spectra, both from the INTERPRET validated database (18). A2, astrocytomas of WHO grade II; OA, oligoastrocytomas of WHO grade II; OD, oligodendrogliomas of WHO grade II; GB, glioblastoma multiforme; ME, metastasis; LY, lymphoma; AB, abscess; MN, meningioma; PN, primitive neuroectodermal tumour; LGG, low-grade glial tumours; AGG, aggressive tumours; WHO, World Health Organization; "n", number of cases used to calculate the mean. Roman numbers stand for the WHO grades of the tumours according to the WHO classification. Colour legend: blue, LGG; red, AGG; brown: WHO grade IV; yellow, AB; grey, MN.

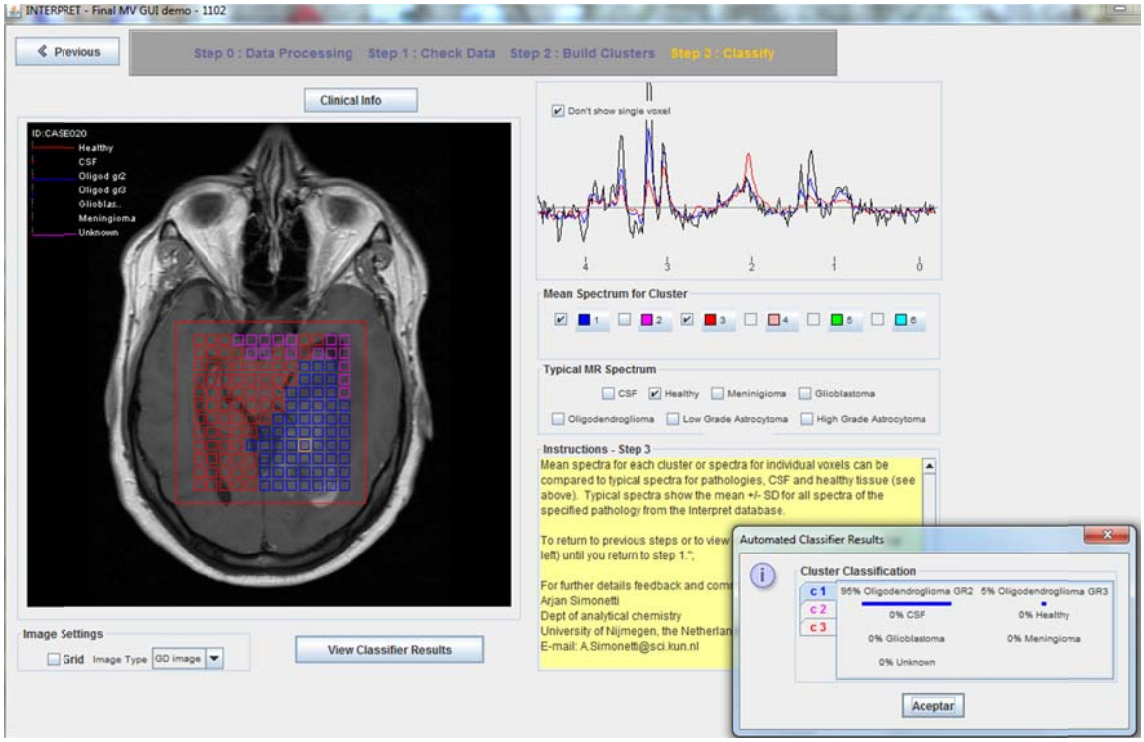

Figure 2. The prototype decision-support system for multivoxel data from Patient I1260, with a diagnosis of OA. The image shows a T1 weighted post gadolinium image on which the results of the clustering algorithm are overlaid. Blue, OD WHO grade II; red, healthy tissue; pink, regions with unspecified diagnosis.

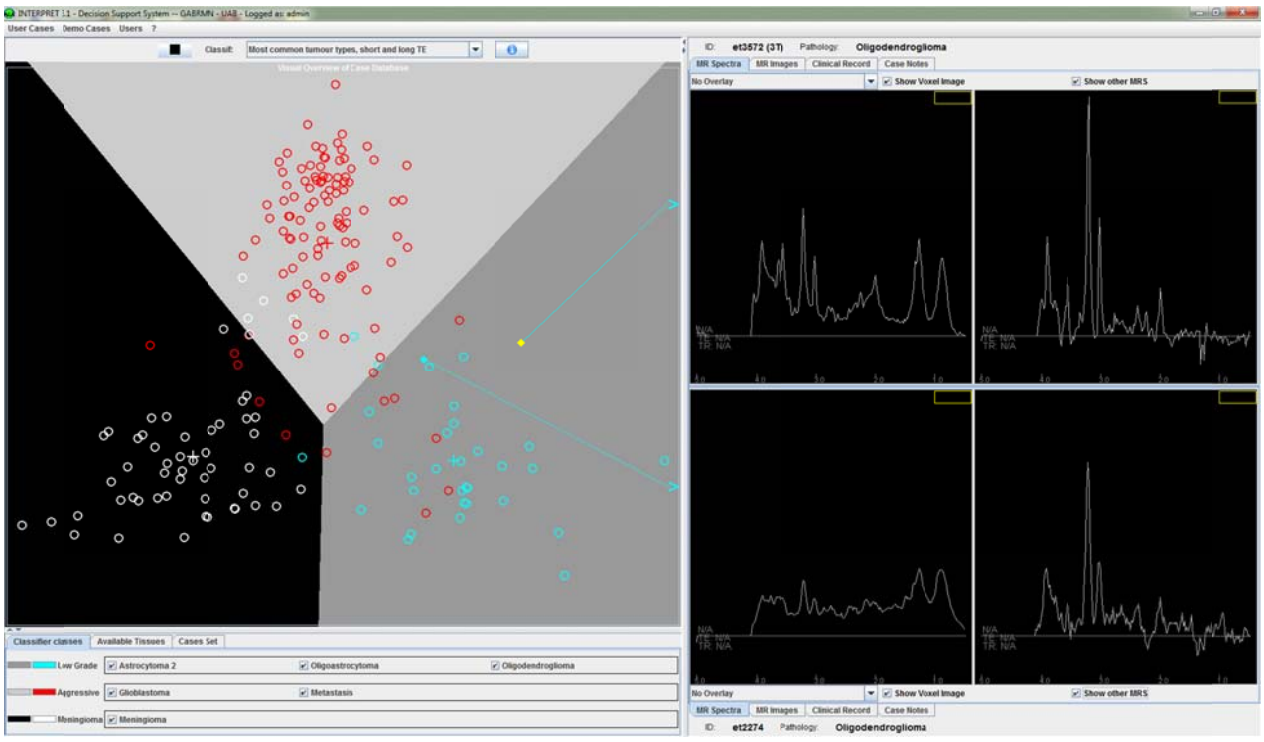

Figure 3. The INTERPRET DSS version 3.1 screen for the "LGG vs. AGG. vs MN" short and long TE classifier. The screen is divided in two main parts, left and right. The overview space of cases in the database is displayed on the

left side, where each case is a coloured circle (see legend on the bottom left). The right part has two panels (top and bottom) for visual inspection of the MRS of individual spectra. In this example, the left panel is subdivided in two subpanels, one for each TE. The top right panels display the short and long TE 3T spectra of et3572, an oligodendroglioma of WHO grade II acquired at the Cambridge partner during eTumour: et3572 is displayed as a yellow symbol in the overview space, and it can be seen that it is correctly classified in the LGG class. The bottom right panels display the short and long TE 1.5 T spectra of case et2274 acquired in Barcelona during the eTumour project (filled blue diamond in the overview space); it is an oligodendroglioma of WHO grade II. With the short TE classifier, the DSS positions et2274 as AGG (not shown), in agreement with another study that also used it (37) and concluded that et2274 was an outlier due to the uncommon mobile lipid pattern (high 0.9 and 1.3 ppm resonances of about the same intensity, disappearing at long TE therefore indicating that these are not necrotic lipids). The long TE classifier and the short and long TE classifier embedded in the DSS however, position et2274 correctly as a LGG. See also reference (39) for compatibility of 3T with 1.5T data classifiers.

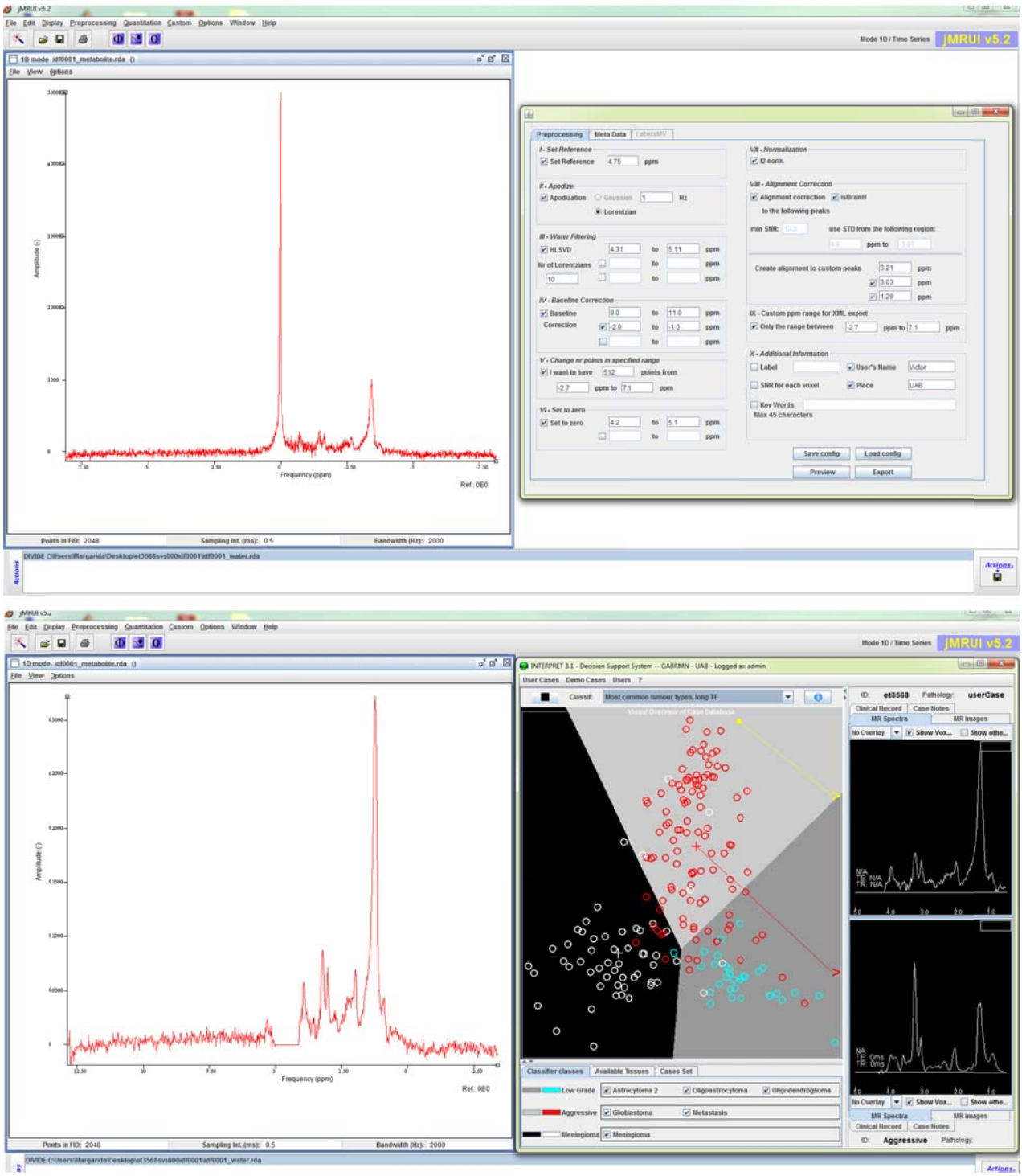

Figure 4. Top, jMRUI with unprocessed spectrum (left panel), long TE case et3568, and jMRUI2XML, in which the INTERPRET processing parameters have been loaded (right panel). Bottom, jMRUI with the processed spectrum according to the INTERPRET processing parameters (left panel) and the DSS in which the same spectrum has been loaded (right panel). In the DSS, the yellow symbol is where this spectrum is positioned by the classifier. The Spectrum can be seen in the top right part of the DSS. It is compared with the mean AGG (bottom right of the DSS). Note that this spectrum corresponds to radiation necrosis, a diagnosis not dealt with by the classifiers handled by the DSS.

	Most common tumour types classifier			Glioblastoma vs. Metastasis classifier	Histopathological diagnosis			
Case number in eTDB	Short TE classifier predicted class	Long TE classifier predicted class	Short+long TE classifier predicted class	Short+Long TE classifier predicted class	Consensus histology by consulting pathologists	Original pathologist	Consulting pathologist A	Consulting pathologist B
et3356	LGG	LGG	LGG	NA	DIFFUSE ASTROCYTOMA 9400/3	DYSEMBRYOPLASTIC NEUROEPITHELIAL TUMOUR 9413/0	DIFFUSE ASTROCYTOMA 9400/3	DIFFUSE ASTROCYTOMA 9400/3
et3460	AGG	AGG	AGG	GBM	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3
et3564	AGG	AGG	AGG	GBM	GLIOBLASTOMA 9440/3	ANAPLASTIC GANGLIOGLIOMA 9505/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3
et3568	AGG	AGG	AGG	NA	OTHER	RADIATION NECROSIS	OTHER	OTHER
et3569	AGG	LGG	AGG	GBM	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3
et3572	LGG	LGG	LGG	NA	OLIGODENDROGLIOMA 9450/3	OLIGODENDROGLIOMA 9450/3	OLIGODENDROGLIOMA 9450/3	OLIGODENDROGLIOMA 9450/3
et3573	AGG	AGG	AGG	NA	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3	GLIOBLASTOMA 9440/3

Table 1. Results of evaluating the seven 3T cases from the eTumour database with the DSS. NA: not applicable. The histopathological diagnoses by the different experts have been included as illustration of the limitations of the gold standard for diagnosing brain tumours.

SUPPLEMENTARY MATERIALS

FOR

Classification of Brain Tumours from MR Spectra: The INTERPRET Collaboration and its Outcomes

By

Margarida Julià-Sapé , John R. Griffiths, A. Rosemary Tate, Franklyn A. Howe, Dionisio Acosta, Geert Postma, Joshua Underwood, Carles Majós, Carles Arús

			Cases contributed			
			iDB			viDB
Contributing centre	**Place**	**Main role**	**Total**	**SV**	**MV**	**SV**
CDP CETIR, Centre Diagnòstic Pedralbes-	Barcelona and Esplugues del Llobregat, Spain	Data contributor	230	230	-	82
IDI, Institut de Diagnòstic per la Imatge-Unitat Bellvitge	L'Hospitalet del Llobregat, Spain	Data contributor	204	204	-	118
SGUL, St George's University of London	London, United Kingdom	Data contributor	159	159	-	75
UMCN, Universitair Medisch Centrum Nijmegen	Nijmegen, The Netherlands	Data contributor	60	46	50	13
UJF, Unité mixte Université Joseph Fourier/INSERM U594	Grenoble, France	Data contributor	70	5	70	-
FLENI, Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia	Buenos Aires, Argentina	Data contributor (Associated partner)	37	37	2	6
MUL, Uniwersytet Medyczny w Lodzi	Lodzi, Poland	Data contributor (Associated partner)	15	15	-	10
UOS, University of Sussex	Brighton, United Kingdom	Database and DSS development, data analysis	-	-	-	-
KUN, Radboud University Nijmegen	Nijmegen, the Netherlands	Data analysis	-	-	-	-
PRAXIM, SARL	Grenoble, France	Commercialisation	-	-	-	-
Siemens AG, Medizinische Technik	Erlangen, Germany	Advisory	-	-	-	-
UAB, Universitat Autònoma de Barcelona	Cerdanyola del Vallès, Spain	Coordination, data managing	-	-	-	-

Table S1. Participating clinical centres as well as roles and number of cases contributed to the databases. Patients scanned at CDP were referred from 6 hospitals in the Barcelona (Spain) area, which contributed clinical data and histology slides. INTERPRET database (iDB), INTERPRET validated database (viDB).

PARAMETER	**STEAM (SHORT TE)**	**PRESS (SHORT TE)**	**PRESS (LONG TE)**
TE	20 ms (20 – 32 ms)	30 -32 ms (30-32 ms)	136 ms (135 – 144 ms)
TR	2000 ms (1600 – 2000 ms)	2000 ms (1600 – 2000 ms)	2000 ms (1600 – 2000 ms)
Volume	4 – 8 cm^3	4 – 8 cm^3	4 – 8 cm^3
Number of averages metabolites	256	192 - 128	192 - 128
Number of averages water	8 to 32	8 to 16	8 to 16
Number of points	512 [Philips]	512 [Philips]	512 [Philips]

	1024 [Siemens] 2048 [GE]	1024 [Siemens] 2048 [GE]	1024 [Siemens] 2048 [GE]
Bandwidth	1000 Hz [Philips] 1000 Hz [Siemens] 2500 Hz [GE]	1000 Hz [Philips] 1000 Hz [Siemens] 2500 Hz [GE]	1000 Hz [Philips] 1000 Hz [Siemens] 2500 Hz [GE]
Dummy scans	4	4	4

Table S2. Consensus acquisition protocols for new data with ranges used for retrospective data accepted into the INTERPRET database. TE and TR ranges used for retrospective data are given in parentheses.

ORDER IN WHICH IT WAS PERFORMED	PROCEDURE
1st	Lineshape correction and zero order phasing using water reference with the Klose method
2nd	0.8 Hz exponential line broadening
3rd	Processing by FFT
4th	Water removal by HLSVD: 5 components removed within ±0.37 ppm of water resonance
5th	Residual water suppression: points at 4.2 to 5.1 ppm set to zero
6th	Linear interpolation to 512 points over 1000 Hz of Siemens and Philips data
7th	Spectrum alignment: maximum of choline peak shifted to 3.21ppm
8th	Normalisation of spectrum to Euclidian norm of peak heights

Table S3. Consensus data processing into the INTERPRET canonical format for spectrum display and analysis.

PROCESS	PARAMETERS	ACCEPT IF
Automatic	Water linewidth (WBW)	WBW < 8Hz
Automatic	S = Maximum metabolite signal in range 0 – 3.4ppm; N = standard deviation noise in range 9 – 11ppm. SNR = S/N	SNR > 10
Manual	Visual inspection by expert spectroscopists. Possible artefacts that cause rejection: high scalp lipids; poor phasing; large baseline artefacts; metabolite peaks of suspect origin	2 experts agree

Table S4. Consensus QC of all spectral data.